Airbnb Super Host

Justin Malunay

Airbnb

What is Airbnb?
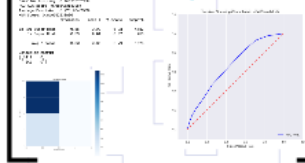- Community marketplace for people to rent their homes, rooms, etc. to others and vice versa
- Founded in 2008 | Guests: 60,000,000 | Cities: 40,000 +
- Castles 1,400 + | Countries: 191 +
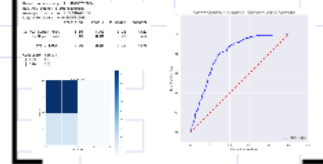- Total revenue in 2004 – 1 million.

Problem Statement:
Super Hosts. What makes a host a "Super" host? I will analyze the features (price, review scores, room type, etc.) that correlate to super hosts and predict how many future hosts there will be using classification (Logistic, KNN, Decision Tree). What is the significance of a super host?
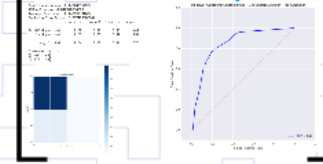
K Nearest Neighbors

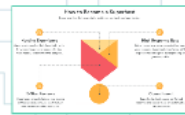## Logistic Regression (L1 penalty)

Decision Tree Classifier

Example :

Super Host Sam

- Bisby = $9,700.00
- Guests = $1,781.40
- Airbnb = $600.00

Super Host Alert? = $1,914.00 minimum
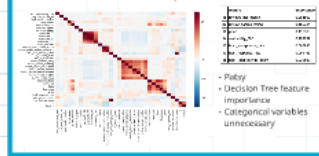
Host Bisby

Booked? = Unknown

What is a "Super Host"?

- Hosted at least 10 trips
- Maintained a 90% response rate or higher
- Received a 5-star review at least 80% of the time you've been reviewed, as long as at least half of the guests who stayed with you left a review
- Completed each of your confirmed reservations without canceling

Feature Importance

- Pixby
- Decision Tree feature importance
- Categorical variables unnecessary

Wait, why are super hosts important?!

How Airbnb makes money:
- Host service fees
  - 3% charge of reservation completion
  - NO profit from hosts
  - Costs go towards processing payments (Paypal, Visa, etc.)
- Guest service fees
  - 6% to 12% depending on size of reservation
- Super Hosts = minimum 10 bookings = $$$

Example:
- Listing = $100.00
- Host = $97.00
- Guest = $106.00 – 112.00

Conclusion / Analysis

- Airbnb's hosts are liquid assets for the business
- No hosts = No guests = No mula baby!
- Super hosts!
- Decision Tree Classifier most successful
  - Predicting future super hosts
- Wish had bookings or sales data
  - Trends and time series analysis
- Projected to hit 900 million

GENERAL ASSEMBLY

Data Info:
- Airbnb Open Data
- San Francisco, CA
- 1 Year (11/01/2015 - 10/30/16)
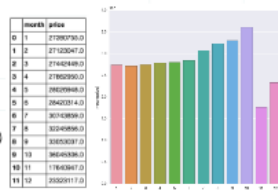- Listings ONLY (**NOT** Bookings !)
- Security/Privacy purposes

Total Hosts: 4647

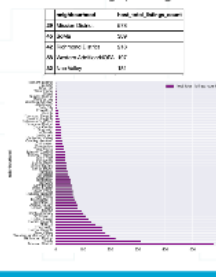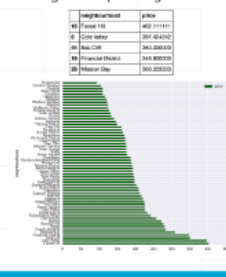|	Super Hosts:	NonHosts:
	723 (~16.0%)	3924 (~84%)
Avg RS:	~97%	~93%
Avg HR:	~98%	~92%

Total price listings per month

Exploratory Data Analysis
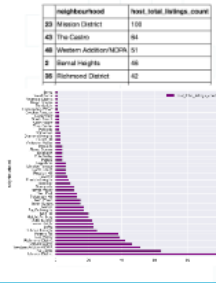
Total host listings per neighborhood

Normal Hosts
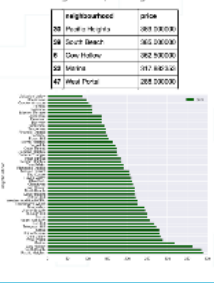Average Price per neighborhood

Exploratory Data Analysis

Total host listings per neighborhood

Super Hosts
Average Price per neighborhood
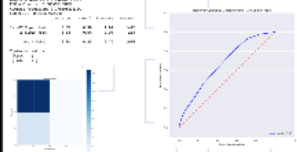
Prezi

Airbnb Super Host

Justin Malunay

Airbnb

What is Airbnb?
- Community marketplace for people to rent their homes, rooms, etc. to others and vice versa
- Founded in 2008 | Guests: 60,000,000 + |Cities: 40,000 +
- Castles 1,400 + | Countries: 191 +
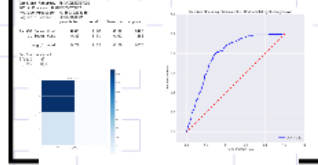- Total revenue in 2009 ~1 Billion

Problem Statement:
Super Hosts. What makes a host a "Super" host? I will analyze the features (price, review scores, room type, etc.) that correlate to super hosts and predict how many future hosts there will be using classification (Logistic, kNN, Decision Tree). What is the significance of a super host?
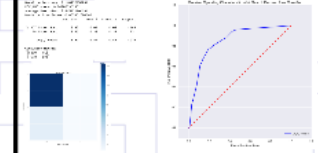
K Nearest Neighbors

Logistic Regression (L1 penalty)
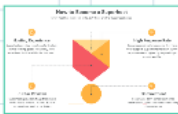
Decision Tree Classifier

Example :

Super Host Sam

Baby = $9,700.00
Guests = 10,000.00
Airbnb = $600.00

Super Host Alert = $1,814.00 minimum
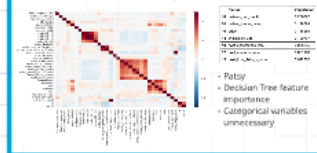
Host Baby

Booked? = Unknown

What is a "Super Host"?

- Hosted at least 10 trips
- Maintained a 90% response rate or higher
- Received a 5-star review at least 80% of the time you've been reviewed, as long as at least half of the guests who stayed with you left a review
- Completed each of your confirmed reservations without canceling

Feature Importance

- Patsy
- Decision Tree feature importance
- Categorical variables unnecessary

Wait, why are super hosts important?!

How Airbnb makes money:
- Host service fees
  - 3% charge of reservation completion
  - NO profit from hosts
  - Costs go towards processing payments (Paypal, Visa, etc.)
- Guest service fees
  - 6% to 12% depending of size of reservation
- Super Hosts = minimum 10 bookings = $$$

Example:
Listing = $100.00
Host = $97.00
Guest = $106.00 - 112.00

Conclusion / Analysis

- Airbnb's hosts are liquid assets for the business
- No hosts = No guests = No mula baby!
- Super hosts!
- Decision Tree Classifier most successful
  - Predicting future super hosts
- Wish had bookings or sales data
  - Trends and time series analysis'
- Projected to hit 900 million

GENERAL ASSEMBLY

Data Info:
- Airbnb Open Data
- San Francisco, CA
- 1 Year (11/01/2015 - 10/30/16)
- Listings ONLY (**NOT** Bookings !)
- Security/Privacy purposes

Total Hosts: 4647

	Super Hosts:	NonHosts:
	723 (~16.0%)	3924 (~84%)
Avg RS:	~97%	~93%
Avg HR:	~98%	~92%

Total price listings per month

Exploratory Data Analysis

Total host listings per neighborhood

Normal Hosts
Average Price per neighborhood

Exploratory Data Analysis

Total host listings per neighborhood

Super Hosts
Average Price per neighborhood

Airbnb

What is Airbnb?
- Community marketplace for people to rent their homes, rooms, etc. to others and vice versa
- Founded in 2008 | Guests: 60,000,000 + | Cities: 40,000 +
- Castles 1,400 + | Countries: 191 +
- Total revenue in 2009 ~ 1 million

Problem Statement:
Super Hosts. What makes a host a "super" host? I will analyze the features (price, review scores, room type, etc.) that correlate to super hosts and predict how many future hosts there will be using classification (Logistc, kNN, Decision Tree). What is the significance of a super host?

What is a "Super Host"?

- Hosted at least 10 trips
- Maintained a 90% response rate or higher
- Received a 5-star review at least 80% of the time you've been reviewed, as long as at least half of the guests who stayed with you left a review
- Completed each of your confirmed reservations without canceling

How to Become a Superhost
Every host has their own style, but these are the Superhost basics.

Hosting Experience
Superhosts develop their hospitality style by welcoming guests frequently. You must host 10 trips within the last year.

High Response Rate
Being responsive shows guests they have your support. Superhosts maintain a 90% response rate or higher by responding to guests quickly.

5-Star Reviews
Superhosts provide listings that inspire enthusiastic reviews. At least 80% of your reviews need to be 5 stars.

Commitment
Superhosts don't cancel confirmed reservations unless there are extenuating circumstances.
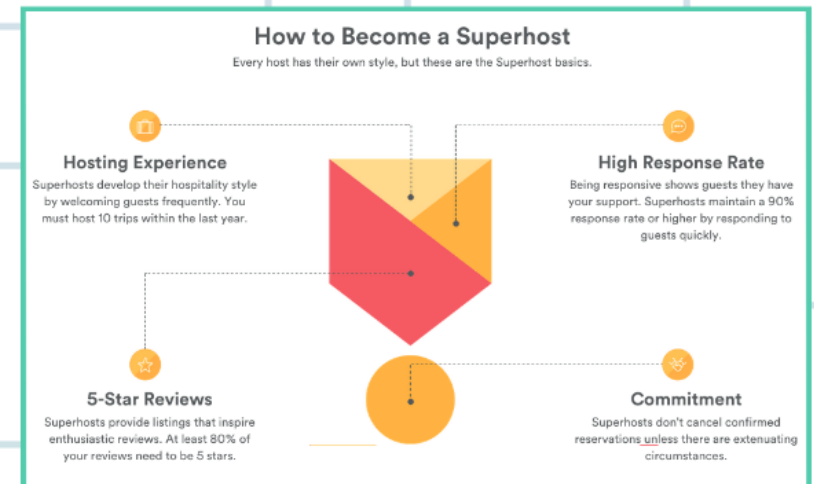
Data Info:
- Airbnb Open Data
- San Francisco, CA
- 1 Year (11/01/2015 - 10/30/16)
- Listings ONLY (**NOT** Bookings !)
- Security/Privacy purposes

Total Hosts: 4647

	Super Hosts:	**NonHosts:**
	723 (~16.0%)	3924 (~84%)
Avg RS:	~97%	~93%
Avg HR:	~98%	~92%

Total price listings per month

	month	price
0	1	27390755.0
1	2	27123047.0
2	3	27442449.0
3	4	27862950.0
4	5	28026948.0
5	6	28420314.0
6	7	30743859.0
7	8	32245856.0
8	9	33053037.0
9	10	36045306.0
10	11	17640947.0
11	12	23323117.0
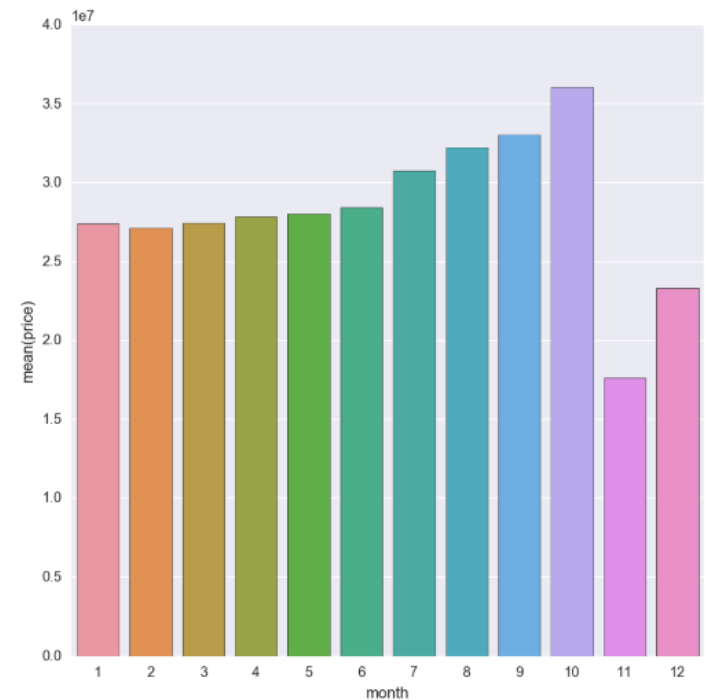
Exploratory Data Analysis

Total host listings per neighborhood

	neighbourhood	host_total_listings_count
29	Mission District	578
45	SoMa	309
42	Richmond District	218
58	Western Addition/NOPA	197
32	Noe Valley	181
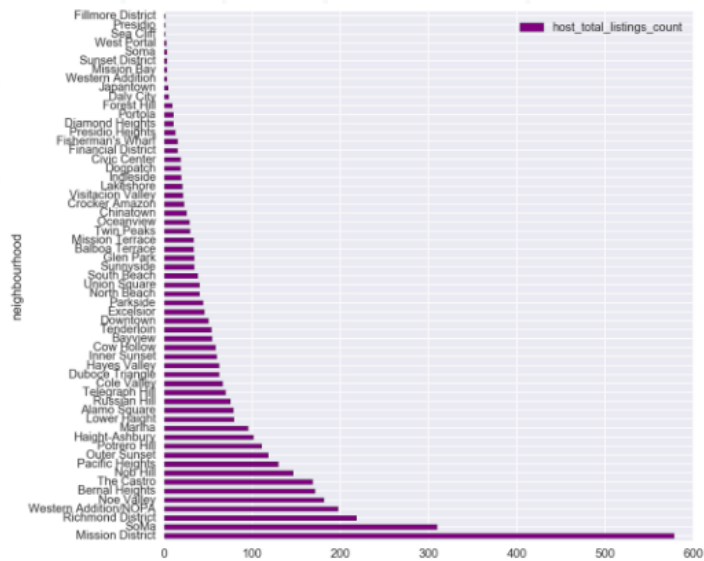
Normal Hosts

Average Price per neighborhood

	neighbourhood	price
18	Forest Hill	402.111111
6	Cole Valley	397.424242
44	Sea Cliff	350.000000
16	Financial District	348.800000
28	Mission Bay	308.333333
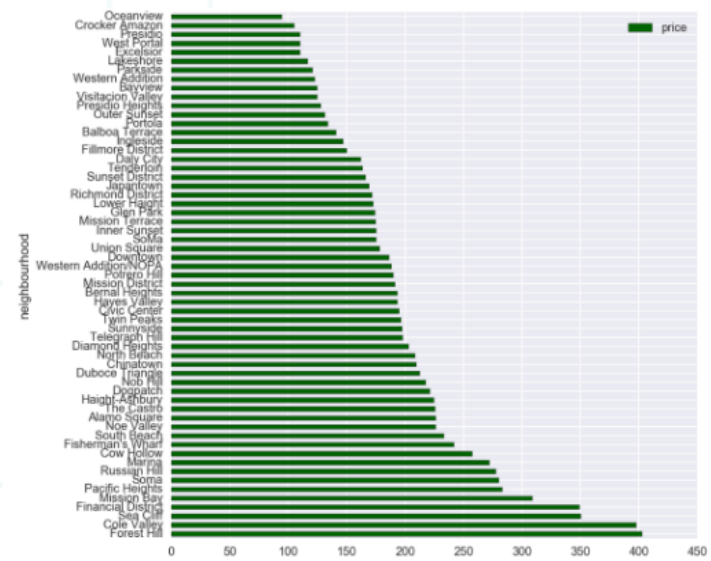
Exploratory Data Analysis

Super Hosts

Total host listings per neighborhood

	neighbourhood	host_total_listings_count
23	Mission District	100
43	The Castro	64
48	Western Addition/NOPA	51
2	Bernal Heights	46
35	Richmond District	42
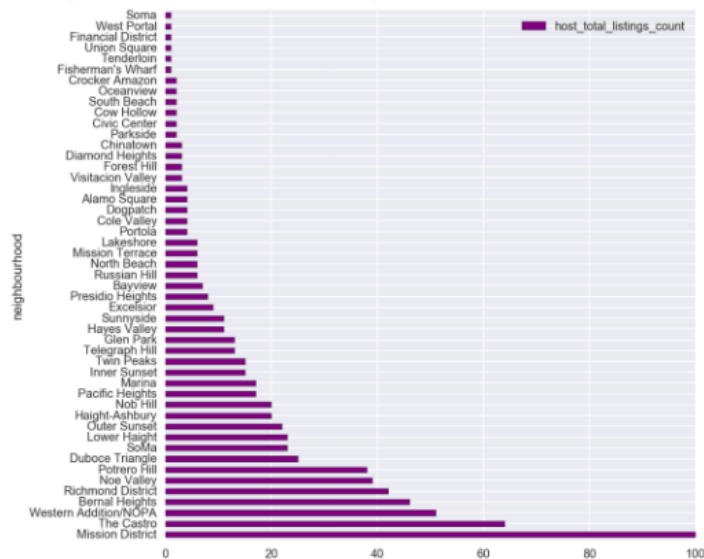
Average Price per neighborhood

	neighbourhood	price
30	Pacific Heights	389.000000
39	South Beach	385.000000
6	Cow Hollow	362.500000
22	Marina	317.882353
47	West Portal	288.000000
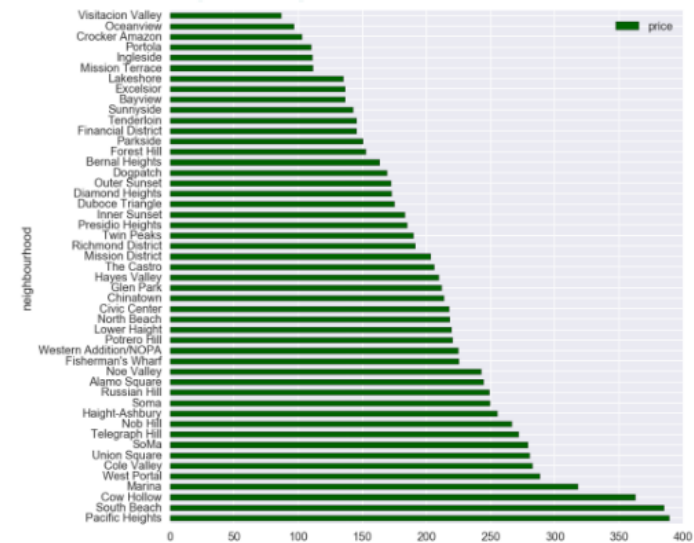
Feature Importance
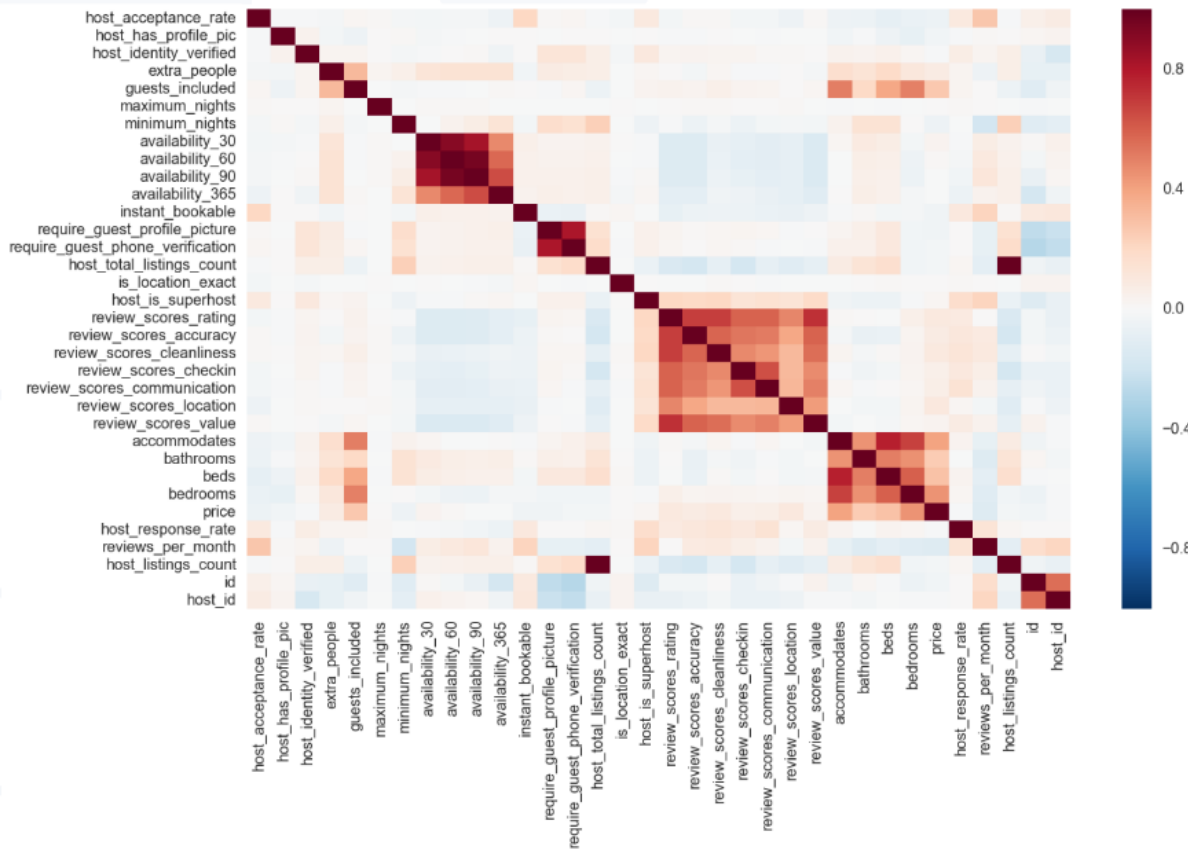
	feature	importance
18	reviews_per_month	0.266897
20	review_scores_rating	0.189784
16	price	0.118969
14	availability_365	0.109797
13	host_acceptance_rate	0.065895
17	host_response_rate	0.061163
15	host_total_listings_count	0.022862

- Patsy
- Decision Tree feature importance
- Categorical variables unnecessary

K Nearest Neighbors

Baseline Accuracy: 0.148235757934
ROC_AUC score: 0.686463133157
Average Precision: 0.276108432224
KNN Score: 0.849552238806

	precision	recall	f1-score	support
Is NOT Super Host	0.85	0.99	0.92	1422
Is Super Host	0.53	0.04	0.07	253
avg / total	0.80	0.85	0.79	1675

Confusion matrix
[[1414   8]
 [244   9]]

Receiver Operating Characteristic - GridSearch kNN

AUC = 0.67
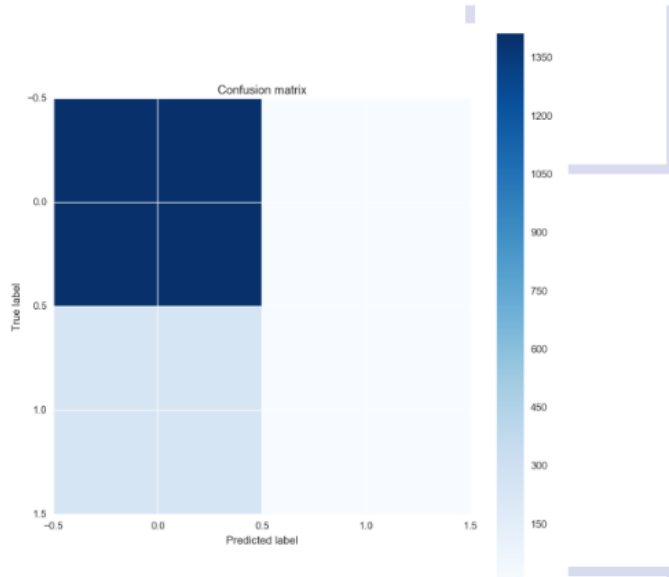
Logistic Regression (L1 penalty)

```
Baseline Accuracy: 0.148235757934
ROC_AUC score: 0.805765076587
Average Precision: 0.35916263188
Logistic Score: 0.848955223881
                        precision    recall    f1-score    support

Is NOT Super Host         0.85        1.00        0.92       1423
    Is Super Host         0.43        0.01        0.02        252

     avg / total          0.79        0.85        0.78       1675

Confusion matrix
[[1419    4]
 [ 249    3]]
```
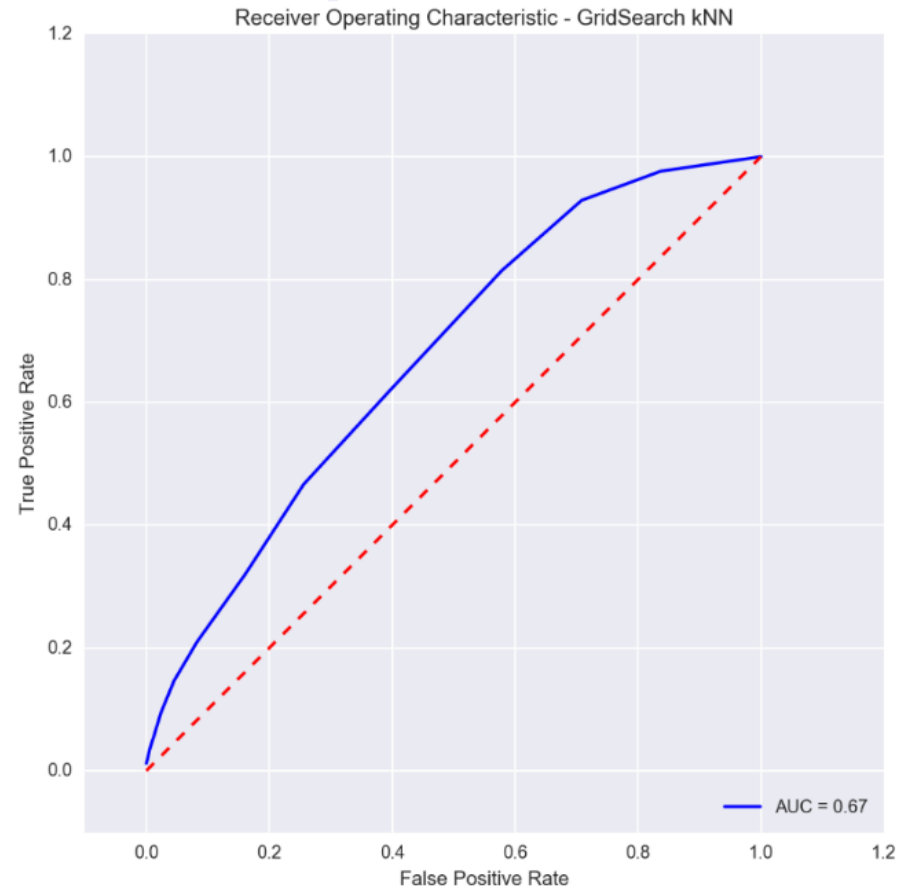

Confusion matrix

Receiver Operating Characteristic - GridSearch Logistic Regression

AUC = 0.80
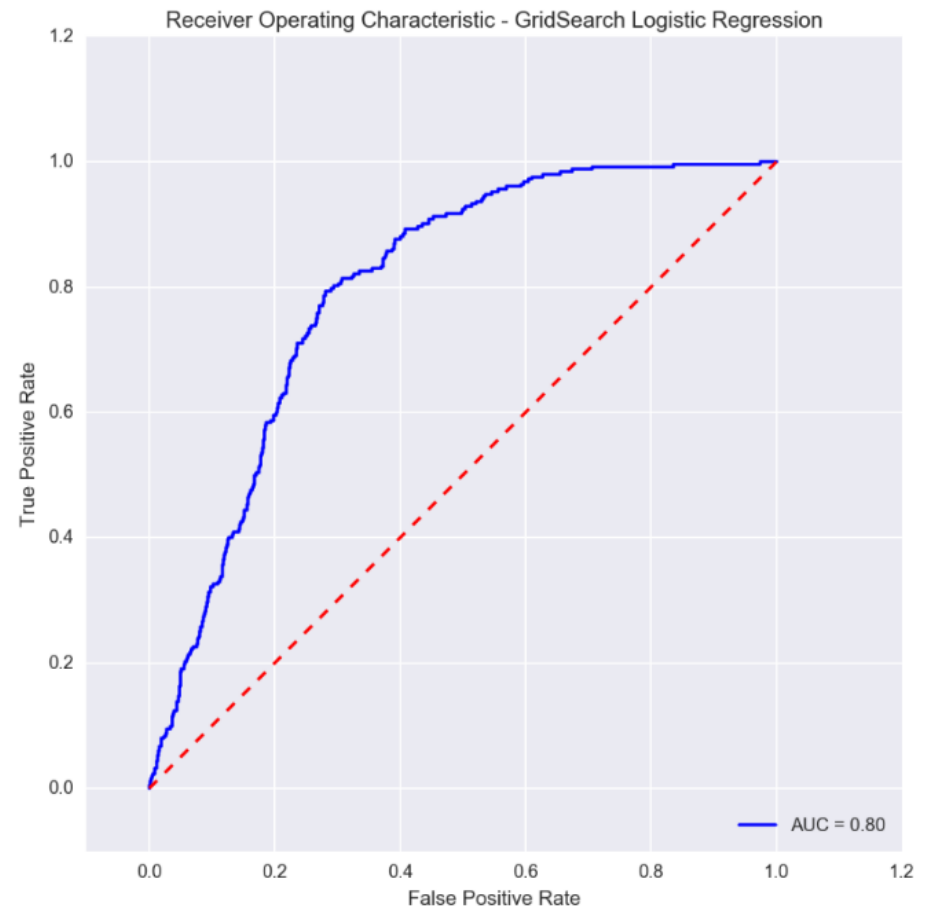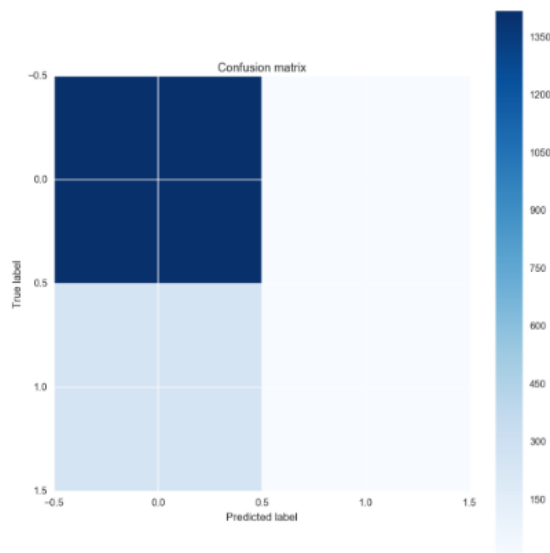
Decision Tree Classifier

Baseline Accuracy: 0.148235757934
ROC_AUC score: 0.840975353362
Average Precision: 0.445122101634
Decision Tree Score: 0.867462686567

	precision	recall	f1-score	support
Is NOT Super Host	0.89	0.97	0.93	1444
Is Super Host	0.55	0.23	0.32	231
avg / total	0.84	0.87	0.84	1675

Confusion matrix
[[1400   44]
 [178   53]]
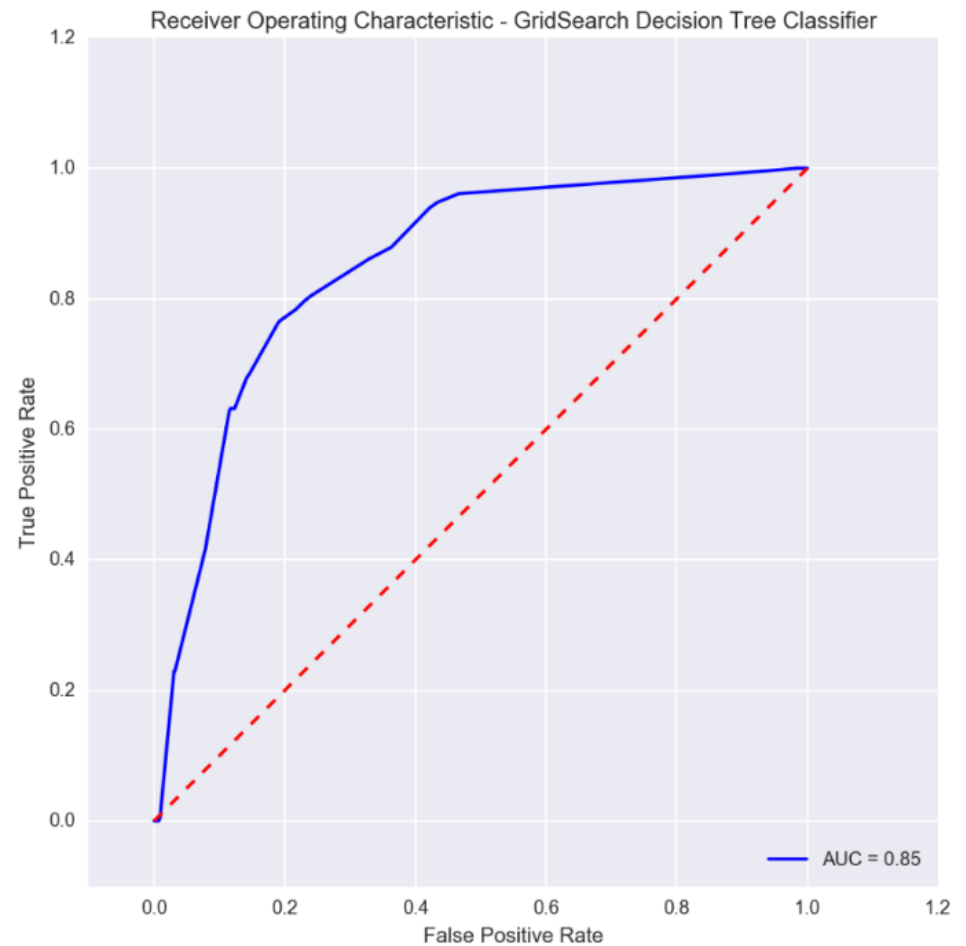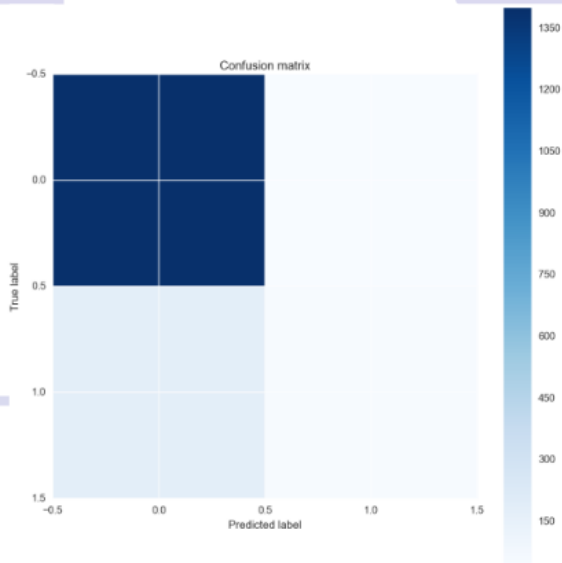
Confusion matrix

Receiver Operating Characteristic - GridSearch Decision Tree Classifier

AUC = 0.85

Wait, why are super hosts important?!

How Airbnb makes money:
- Host service fees
  - 3% charge of reservation completion
  - NO profit from hosts
  - Costs go towards processing payments (Paypal, Visa, etc.)
- Guest service fees
  - 6% to 12% depending of size of reservation
- Super Hosts = minimun 10 bookings = $$$

Example:
Listing = $100.00
Host = $97.00
Guest = $106.00 - 112.00

Example :

Super Host Sam

	host_name	neighbourhood	price	host_total_listings_count
2511	Sam	Mission District	1690.0	1.0

Sam = $1,639.30
Guests = $1,791.40
Airbnb = $101.40

Super Host Alert! = $1,014.00 minimum

Bixby = $9,700.00
Guests = 10,600.00
Airbnb = $600.00

Host Bixby

Booked? = Unknown

	host_name	neighbourhood	price	host_total_listings_count
84	Bixby	Cole Valley	10000.0	1.0

Prezi

Conclusion / Analysis

- Airbnb's hosts are liquid assets for the business
- No hosts = No guests = No mula baby!
- Super hosts!
- Decision Tree Classifer most successful
  - Predicting future super hosts
- Wish had bookings or sales data
  - Trends and time series analysis'
- Projected to hit 900 million
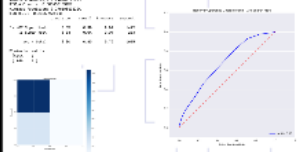
Airbnb Super Host

Justin Malunay

Airbnb

What is Airbnb?
- Community marketplace for people to rent their homes, rooms, etc. to others and vice versa
- Founded in 2008 | Guests: 60,000,000 + |Cities: 40,000 +
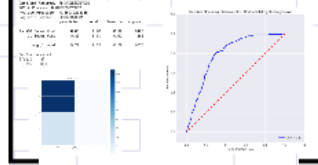- Cities: 1,400 + | Countries: 191 +
- Total revenue in 2009 ~1 Billion

Problem Statement:
Super Hosts. What makes a host a "Super" host? I will analyze the features (price, review scores, room type, etc.) that correlate to super hosts and predict how many future hosts there will be using classification (Logistic, kNN, Decision Tree). What is the significance of a super host?
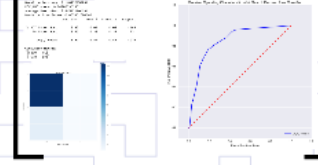
K Nearest Neighbors

Logistic Regression (L1 penalty)
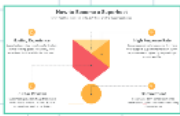
Decision Tree Classifier

Example :

Super Host Sam

Bixby = $9,766.00
Earn = $1,629.20
Guests = 10,000.00
Guests = $1,731.40
Airbnb = $901.00
Airbnb = $869.00

Super Host Alert = $1,814.00 minimum

Host Bixby

Booked? = Unknown

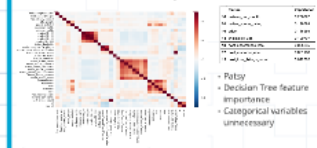| host_name	neighborhood		host_total_listings_count
Sam	Cole Valley	10000.00	10

What is a "Super Host"?

- Hosted at least 10 trips
- Maintained a 90% response rate or higher
- Received a 5-star review at least 80% of the time you've been reviewed, as long as at least half of the guests who stayed with you left a review
- Completed each of your confirmed reservations without canceling

Feature Importance

- Patsy
- Decision Tree feature importance
- Categorical variables unnecessary

Wait, why are super hosts important?!

How Airbnb makes money:
- Host service fees
  - 3% charge of reservation completion
  - NO profit from hosts
  - Costs go towards processing payments (Paypal, Visa, etc.)
- Guest service fees
  - 6% to 12% depending of size of reservation
- Super Hosts = minimum 10 bookings = $$$

Example:
Listing = $100.00
Host = $97.00
Guest = $106.00 - 112.00

Conclusion / Analysis

- Airbnb's hosts are liquid assets for the business
- No hosts = No guests = No mula baby!
- Super hosts!
- Decision Tree Classifier most successful
  - Predicting future super hosts
- Wish had bookings or sales data
  - Trends and time series analysis'
- Projected to hit 900 million

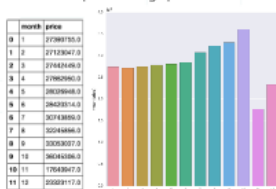**GENERAL ASSEMBLY**

Data Info:
- Airbnb Open Data
- San Francisco, CA
- 1 Year (11/01/2015 - 10/30/16)
- Listings ONLY (**NOT** Bookings !)
- Security/Privacy purposes

Total Hosts: 4647

	Super Hosts:	NonHosts:
	723 (~16.0%)	3924 (~84%)
Avg RS:	~97%	~93%
Avg HR:	~98%	~92%

Total price listings per month

	month	price
0	1	27398755.0
1	2	27123047.0
2	3	27442449.0
3	4	27985290.0
4	5	28028948.0
5	6	28420314.0
6	7	30743889.0
7	8	30246866.0
8	9	30053097.0
9	10	36046586.0
10	11	17643947.0
11	12	35023117.0

Exploratory Data Analysis

Total host listings per neighborhood

	neighbourhood	host_total_listings_count
25	Mission District	475
6	Cole	305
47	Financial District	210
18	Western Addition/NOPA	187
33	Noe Valley	183

Normal Hosts
Average Price per neighborhood

	neighbourhood	price
7	Potrero Hill	997.490940
3	Cole Valley	987.434047
44	Sea Cliff	966.000000
16	Financial District	946.000000
25	Mission Bay	906.000000

Exploratory Data Analysis

Total host listings per neighborhood

	neighbourhood	host_total_listings_count
19	Pacific Heights	100
45	The Castro	64
48	Western Addition/NOPA	51
2	Bernal Heights	46
39	Richmond District	42

Super Hosts
Average Price per neighborhood

	neighbourhood	price
19	Pacific Heights	595.000000
39	South Beach	585.000000
5	Dow Hollow	566.600000
22	Marina	547.000255
47	West Portal	544.000000

Prezi